Scaling Prospects for Ultimate Nanotransistors

ONR Grant # N00014-04-1-0542

SUNY RF Award #32804

Final Technical Report

Principal Investigator: Prof. Konstantin K. Likharev
Department of Physics and Astronomy
Stony Brook University
Stony Brook, NY 11794-3800

Phone: 516-632-8159
Fax: 516-632-8774
E-mail: klikharev@notes.cc.sunysb.edu

Performance Period: May 1, 2004 – February 28, 2005

May 2005

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
31/05/2005	Final Technical Report	From 1 May 2004 to 28 Feb 2005

4. TITLE AND SUBTITLE

Scaling Prospects for Ultimate Nanotransistors

5a. CONTRACT NUMBER

N00014-04-1-0542

5b. GRANT NUMBER

SUNY RF #32804

5c. PROGRAM ELEMENT NUMBER

n/a

6. AUTHOR(S)

Likharev, Konstantin K.

5d. PROJECT NUMBER

SUNY RF #1040559

5e. TASK NUMBER

SUNY RF #1

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Research Foundation of SUNY at Stony Brook
West 5510 Frank Melville Memorial Library
Stony Brook, NY 11794-3362

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approve for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Advanced semiconductor field-effect transistors (FET) scaled into the sub-10-nm gate length range, are sometimes considered the main candidates for future nanoelectronics even beyond the long-term horizon of the International Technology Roadmap for Semiconductors. In this project, prospects of FET scaling have been evaluated in more detail, than has been done before. In particular, we have calculated the source-drain I-V curves, subthreshold characteristics, voltage gain, and power consumption of sub-10-nm double-gate silicon MOSFETs using the self-consistent solution of quasi-2D Schrödinger and 2D Poisson equations. Most importantly, the sensitivity of transistor characteristics (in particular, the gate voltage threshold) to variations of the structure dimensions, have been evaluated in detail. The results has shown that this sensitivity, strongly affecting the fabrication facilities costs, sets the ultimate limits for CMOS technology scaling. Based on our results, this limit is close to 10 nm gate length for single-gate and 8 nm for double-gate transistors.

15. SUBJECT TERMS

Microelectronics, nanoelectronics, FET, MOSFET, CMOS, voltage gain, power consumption, parameter sensitivity, fabrication costs

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
			UU	10	19b. TELEPHONE NUMBER *(Include area code)*

THE RESEARCH FOUNDATION
The State University of New York

Stony Brook University

Office of Sponsored Programs

Stony Brook, New York
11794-3362

Phone 631 632 4402/9949
Fax 631 632 6963

Office of Naval Research
Ballston Centre Tower One
800 North Quincy Street
Arlington, VA 22217-5660

Re: N00014-04-1-0542, PI: Konstantin K. Likharev

Dear Colleague:

Enclosed please find attached the final progress report for the above referenced award.

Any questions regarding this submission should be directed to Jennifer Hemmerick at jhemmerick@notes.cc.sunysb.edu or by telephone at (631) 632-9079.

Sincerely,

Susan Gasparo 6/24/05
Susan Gasparo
Sponsored Programs Coordinator

CC:
Office of Naval Research Regional Office4 Boston
495 Summer Street Room 627
Boston, MA 02210-2109

Defense Technical Information Center
8725 John J. Klingman Road Ste 0944
Fort Belvoir, VA 22060-6218

Naval Research Laboratory
ATTN: Code 5227
4555 Overlook Avenue SW
Washington DC20375-5320

1. Executive Summary

Advanced semiconductor field-effect transistors (FET), scaled into the sub-10-nm gate length range, are sometimes considered the main candidates for future nanoelectronics even beyond the long-term horizon of the International Technology Roadmap for Semiconductors. In this project, long-term prospects of FET scaling have been evaluated in more detail than before. In particular, we have calculated the source-drain *I-V* curves, subthreshold characteristics, voltage gain, and power consumption of sub-10-nm double-gate silicon MOSFETs using the self-consistent solution of quasi-2D Schrödinger and 2D Poisson equations. Most importantly, the sensitivity of transistor characteristics (in particular, the gate voltage threshold) to variations of the structure dimensions, have been evaluated in detail. The results has shown that this sensitivity, strongly affecting the fabrication facilities costs, sets the ultimate limits for CMOS technology scaling. Based on our results, this limit is close to 10 nm gate length for single-gate and 8 nm for double-gate transistors. The further continuation of the Moore Law development of microelectronics will probably require transfer to integrated circuits based on CMOS/nanodevice hybrids.

2. Model

Figure 1 shows our models for the two versions of SOI MOSFETs. In both transistor options, the gate is assumed to be aligned with the undoped section of the SOI channel that is connected to bulk source and drain via their thin, doped extensions. In the sub-10-nm gate length range, channel doping is unnecessary [2, 3]. Estimates show [8] that this allows us to neglect electron scattering in the undoped part of the channel. On the other hand, the doping of source and drain (including their thin extensions) should be extremely high ($N_D \gtrsim 10^{20}$ cm^{-3}) to ensure device reproducibility [4, 7, 8].

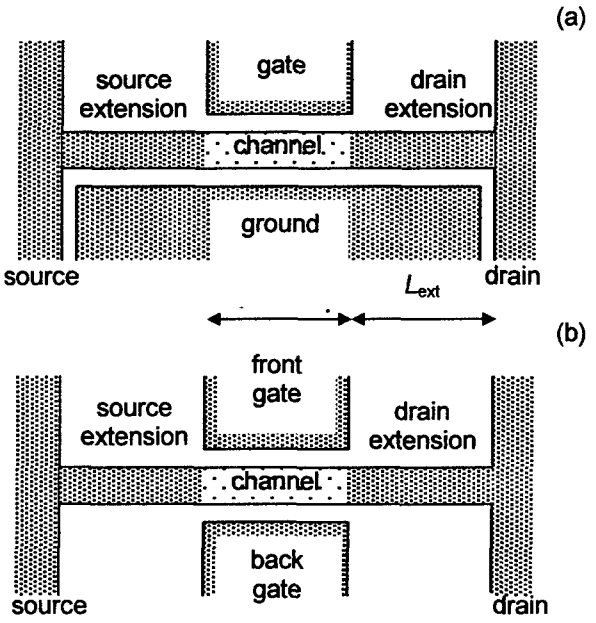

Figure 1. The models of (a) single-gate and (b) double-gate nanoscale SOI MOSFETs, used in this study. All the results shown below have been calculated for the doped extension length L_{ext} = 10 nm, and the electrode doping level $N_D = 3\times10^{20}$ cm^{-3}. (While the former parameter does not affect results too much, the choice of the latter one in the sub-10-nm range of L is not too broad [4]).

In the single-gate version of the device (Fig. 1a), the ground electrode is extended under almost the entire extension region, and its electric potential is equal to that of the source. On the other

hand, in the double-gate transistor (Fig. 1b) the back gate is electrically connected to the front gate, and has the same length L. These geometries is typical for most methods of SOI MOSFET fabrication.

In this size range, a full quantum-mechanical treatment of electron transfer, in particular of the electron confinement in the direction across the ultrathin channel, and the source-to-drain tunneling in the direction along the channel, is a must. Indeed, the latter effect provides one of the major limitations for transistor scaling [4, 8, 9]. Another major effect is a gradual loss of electrostatic control of the channel potential as the gate length L is decreased. Just as in Refs. 7, 8, we address these problems by the self-consistent solution of the quasi-2D Schrödinger equation and 2D Poisson equation. (We have shown [8] that this approach gives results similar to those of the more fashionable NEGF formalism.)

3. Main Results

A. Potential Profile

Figure 2 shows typical distributions of the conduction band edge along the transistor length (partly including thin extensions of the source and drain). Since the two MOSFET species under analysis are depleted at different gate voltages V_G, we have plotted them for approximately the same value of the drain current to provide a more fair comparison.

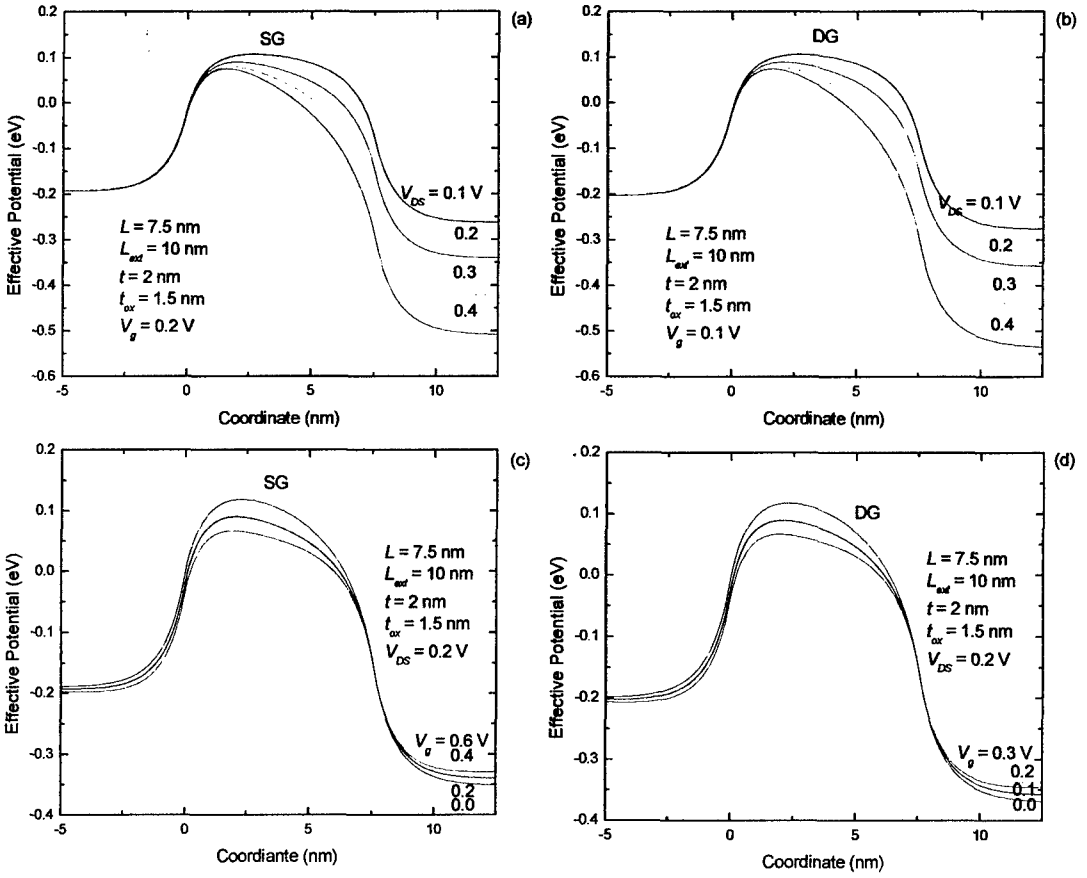

Figure 2. Effective potential profile for: (a, c) single-gate and (b) double-gate MOSFETs with gate length L = 7.5 nm for the same drain current. Panels (a) and (b) show the results of change of the drain-source voltage, while panels (c) and (d) illustrate the effect of gate voltage variation. The origin of coordinate x (along the transistor length) is at the interface between the doped

source extension and undoped channel (of the same thickness t). Transistor parameters are shown at the panels. (Here and below t_{ox} is the gate oxide thicknesses.)

One can see that the potential profiles in both cases are very similar, with the exception that the sensitivity of the potential peak (that essentially controls the drain current) to gate voltage is approximately twice stronger for the double-gate transistor. For negative V_g (depleting the channel) this is expected. In the single-gate device the electrostatic potential changes almost linearly between the gate and the ground electrodes, with half V_g being in the center of the channel. In the double-gate transistor, the potential is approximately constant in the direction across the channel. However, we had not expected that the same relation would hold even in the case of positive V_g, when electrons in the channel screen the gate potential substantially.

B. Source-Drain I-V Curves

Figure 3 shows typical families of I-V curves for both transistors. In both cases, the reduction of the gate length in the single-gate transistors leads to an almost similar suppression of current saturation at larger drain-source voltages. (For the used values of t and t_{ox}, this deterioration is due to, in almost equal parts, to the electrostatics degradation and the onset of source-to-drain tunneling [8, 9].) Again, the main difference between the two MOSFET species is that it takes (approximately) twice more gate voltage to change the current in the single-gate transistor by a certain amount, than in its double-gate counterpart.

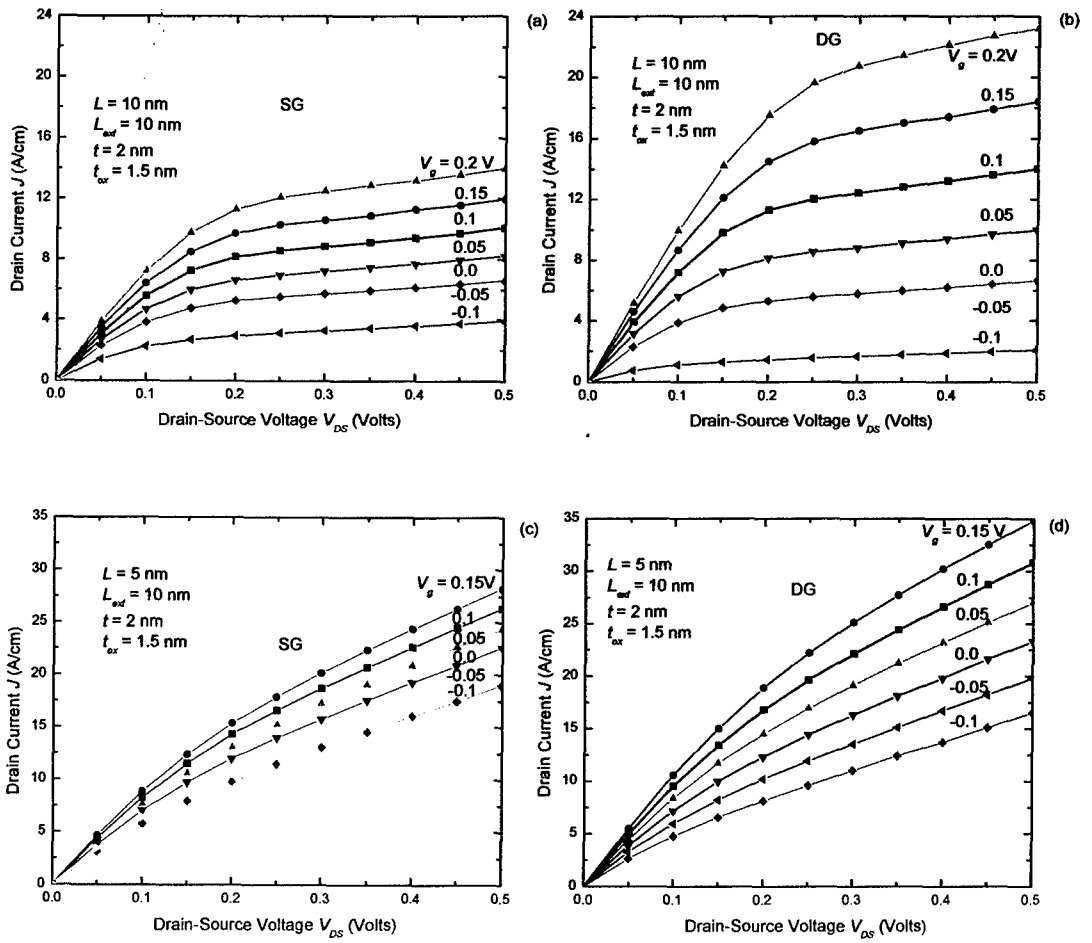

Figure 3. Source-drain I-V curves of (a, c) single-gate and (b, d) double-gate transistors with channel length (a, b) L = 10 nm and (c, d) L = 5 nm, for several values of the gate voltage V_g.

C. Subthreshold Curves

The same effect can be seen in Fig. 4 showing typical subthreshold curves of both devices. Even in relatively long single-gate devices, the slope is at least twice less than the perfect value of 60 mV/decade.

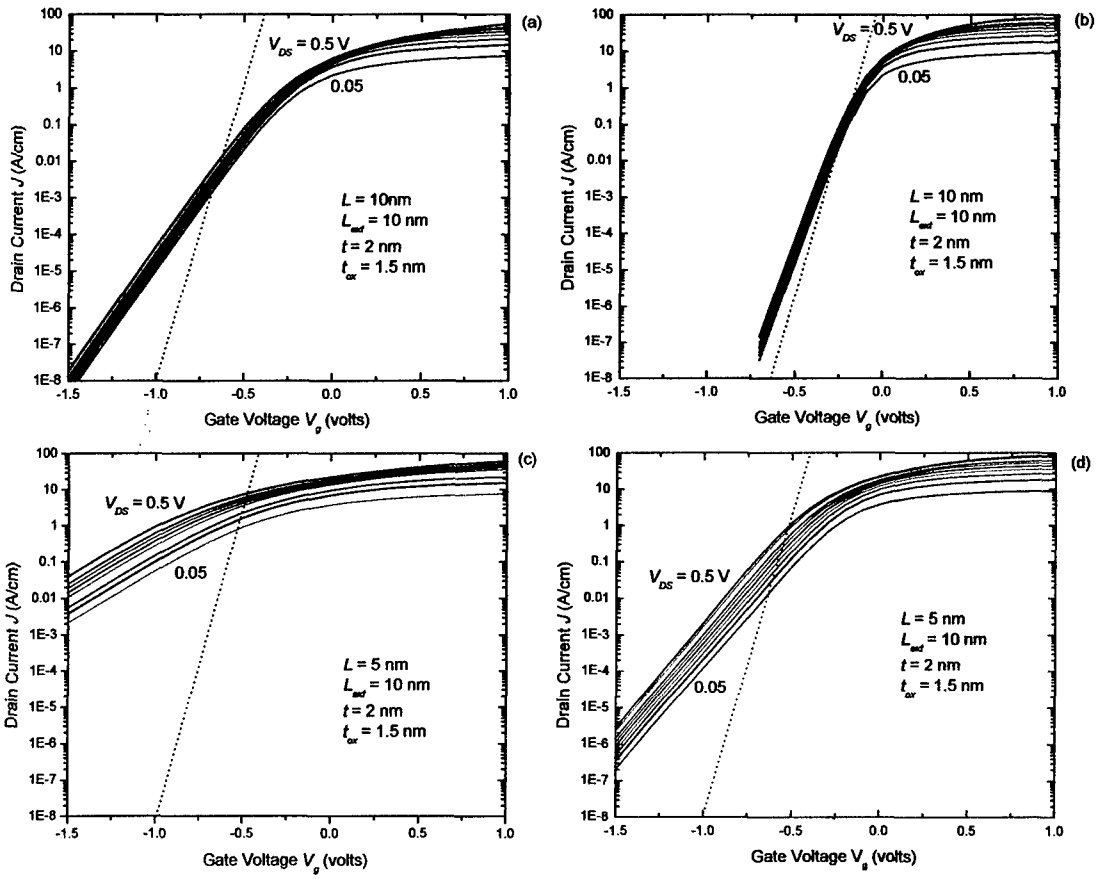

Figure 4. Subthreshold curves of (a, c) single-gate and (b, d) double-gate transistors with channel length (a, b) L = 10 nm and (c, d) L = 5 nm, for ten values of the drain-source voltage in each case (with 50-mV steps). The dashed lines show the perfect slope (~60 mV/decade).

D. Voltage Gain

The degradation of the gate control of both ON and OFF currents can be characterized by a single parameter, voltage gain, i.e. the partial derivative $G_V \equiv \partial V / \partial V_g$ taken at a fixed drain current density. In good MOSFETs, $G_V \to \infty$ at saturation, so this is not a very popular engineering figure-of-merit. However, as the transistor degrades, the voltage gain becomes an important characteristic, since digital logic circuits fundamentally require $G_V > 1$ for their operation. Figure 5 shows G_V as a function of the drain current, at fixed source-drain voltage. (Just as in the case of potential profiles, this comparison is more fair than it would be if the gain were plotted against the gate voltage V_G.) One can see that again, in single-gate transistors the gain is always approximately half that of the double-gate transistor. This is why when G_V falls with decreasing L due to device degradation, the device usefulness boundary (G_V = 1) is reached by the single-gate transistor first (at $L \approx 3$ nm), while the double-gate device can provide gain even at L = 2 nm.

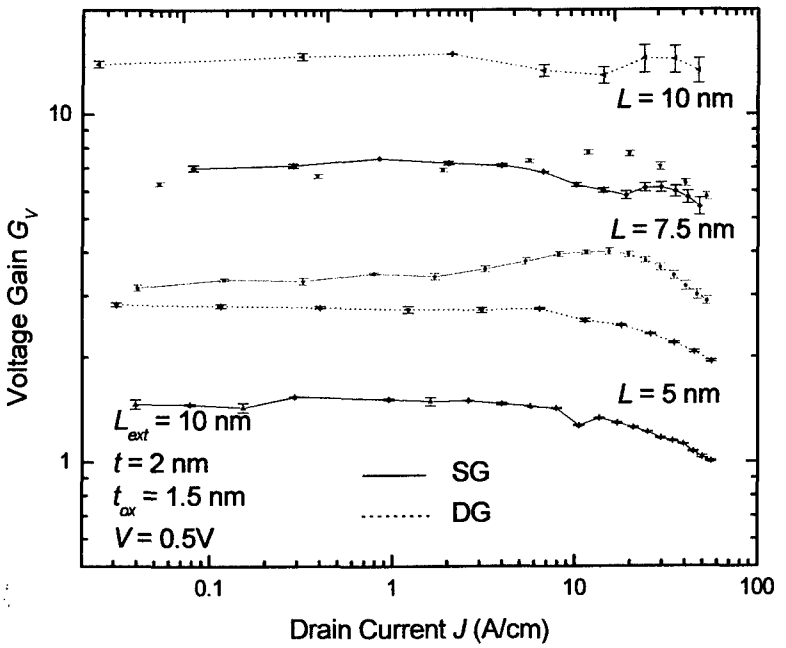

Figure 5. Voltage gain $G_V \equiv \partial V_{DS}/\partial V_g|_{J=\text{const}}$ as a function of drain current density, for several values of gate length.

E. Power

Well before the MOSFETs are scaled down to the gain loss point, they acquire two features unfavorable for applications. The first one is a growth of power, even at optimally selected gate workfunction and power supply voltage V_{DD} (which in CMOS circuits also gives the signal swing). In order to analyze this effect we have used a simple model for the total power in CMOS circuits [10] (it is also described in detail in Ref. 8). Although that model is approximate, we believe it captures the basic interplay between the static and dynamic power. The main advantage of the model is that after gate workfunction optimization, the specific power (per unit channel width) depends on just three parameters: the power supply voltage V_{DD}, ON current density J_{ON}, and a "switching activity parameter" λ, typically of the order of 10^{-2}. (The results are insensitive to λ, unless this parameter is impracticably large.)

Figure 6 shows the total power and its components as functions of V_{DD}, for a typical values of λ and J_{ON} [1]. Static power decreases with V_{DD}, because larger voltage swing allows enables the transistors to be closed better in the depletion region (Fig. 4). On the other hand, dynamic power grows with V_{DD} (in the normalization used in Fig. 6, linearly). As a result, total power as a function of V_{DD} has a minimum [8-10].

When carrying out this procedure for the single-gate MOSFETs, we have run into the following new problem. If such a transistor is small enough, power optimization brings us to values of the signal swing so large that when the transistor is nominally closed ($V_g = 0$, $V_{DS} = V_{DD}$), the valence band edge at some internal point of the channel becomes higher than the conduction band edge in the drain (Fig. 6b), even taking into account the bandgap broadening due to quantum confinement. In channels so short, this band overlap immediately leads to intensive Zener tunneling [11] from internal points of the channel to drain, forming holes in the channel near its interface with the source (see the arrow in the inset of Fig. 6b).

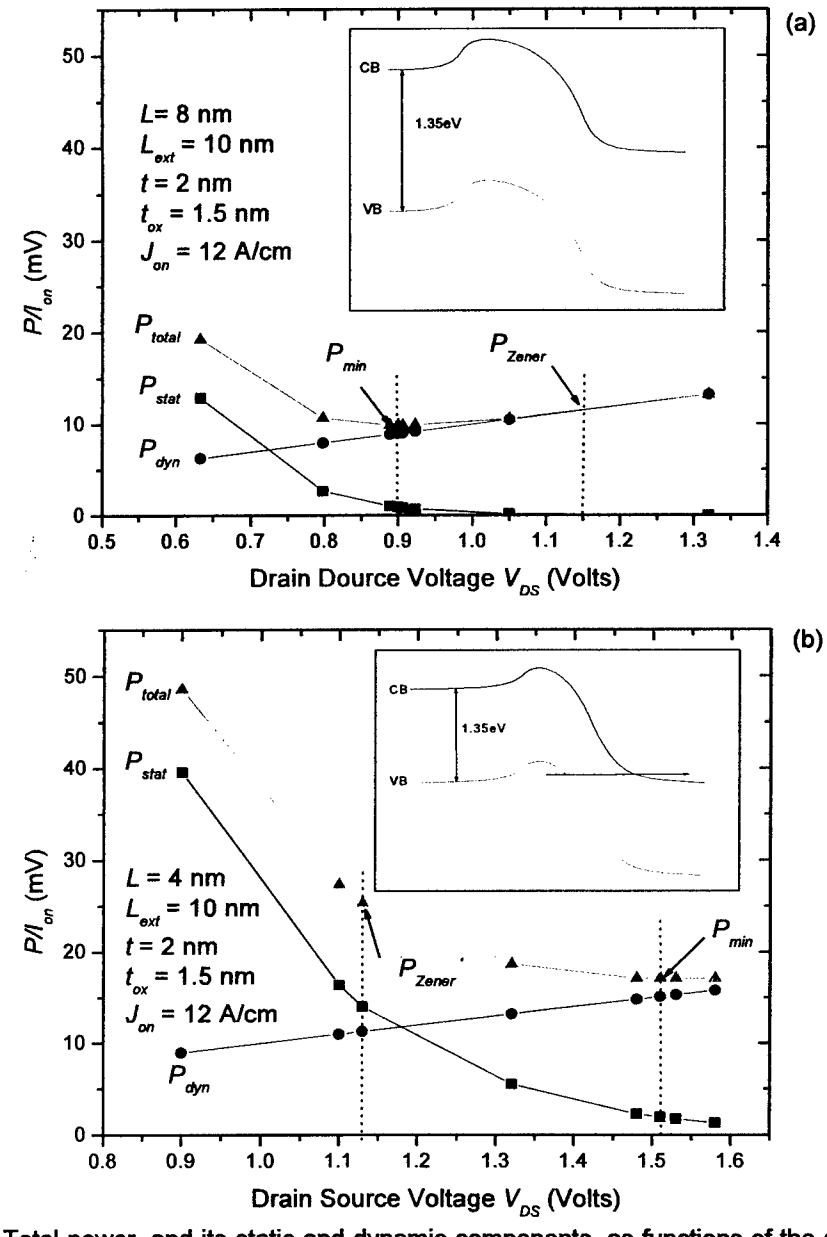

Figure 6. Total power, and its static and dynamic components, as functions of the source-drain voltage for (a) L = 8 nm and (b) L = 4 nm. The insets show the band edge diagram for (a) V_{DS} = 0.9 volt and (b) V_{DS} = 1.52 volt. Dashed lines show the point of minimum power without the account for the Zener effect (P_{min}), and the point where this effect starts (P_{Zener}).

This loss may be only compensated by electron-hole recombination in the channel. Though this effect is unaccounted for in our theory, estimates show it is far insufficient to compensate the Zener tunneling, so that a positively-charged "pocket" of holes would appear inside the channel, lowering the electrostatic potential and preventing the device from shutting down.

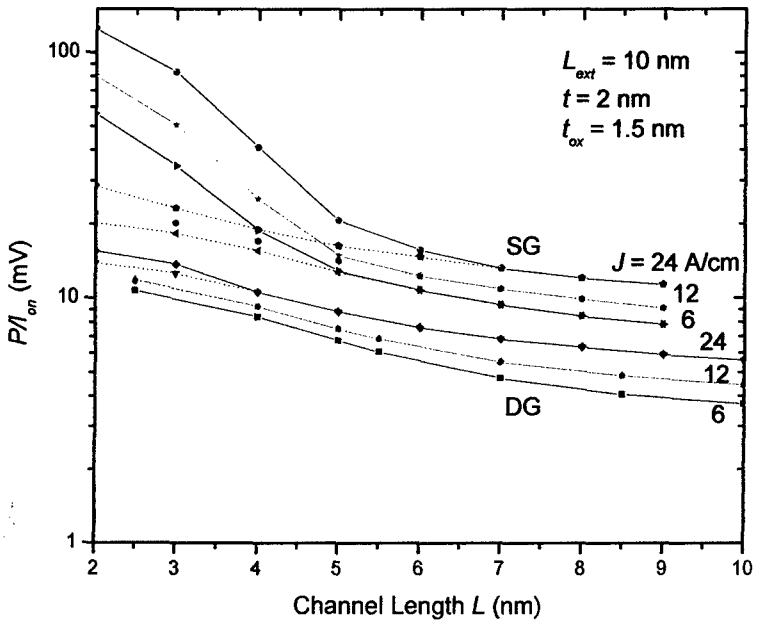

Figure 7. Minimum total power as a function of channel length. Dotted lines show the results of a calculation which neglects the Zener tunneling effect.

In order to account for this effect, we have set an upper bound on V_{DD}, limiting it by the point of band edge crossing (i.e., the Zener tunneling onset). This immediately has resulted is an increase of the total power minimum. Figure 7 shows the resulting minimum power as a function of transistor length, for several values of ON current, corresponding to long-term ITRS goals [1]. One can see that in the single-gate transistors the Zener tunneling effect begins at $L \leq 6$ nm, while in double-gate devices it is not essential in all the gate length of interest (thus justifying our previous results [7, 8]).

F. Parameter Sensitivity

Another scaling effect, which may have even larger negative practical impact than the power growth, is the exponentially growing sensitivity of transistor characteristics (most importantly, the threshold value V_t of gate voltage) to minute variations of device dimensions [4, 8, 9]. Figure 8 shows the threshold voltage "roll-off" (i.e., the difference $|V_t(L) - V_t(\infty)|$, as a function of gate length. The results show that for the single-gate transistors this "sensitivity crisis" starts two nanometers or so earlier than in their double-gate counterparts.

4. Conclusions

Our calculations have shown that single-gate SOI MOSFETs differ from double-gate SOI devices (with similar parameters) by twice weaker control of the current by the gate voltage. It is interesting that this relation of the two devices, evident in the depletion range, is also extended, almost exactly, over the ON state range.

By itself, this transconductance (and hence gain) loss might be not very detrimental, taking into account the simpler fabrication technology and also the fact that the internal logic delay of two transistors is comparable (due to the twice larger input capacitance of double-gate transistors). However, single-gate geometry exacerbates two main fundamental problems of the "ultimate" (sub-10-nm) MOSFET scaling: the exponential growth of power consumption and sensitivity to fabrication uncertainties. These factors will play the decisive role in for eventual transfer of the CMOL industry to double-gate SOI devices and then to CMOS/nanodevice hybrids [4, 11].

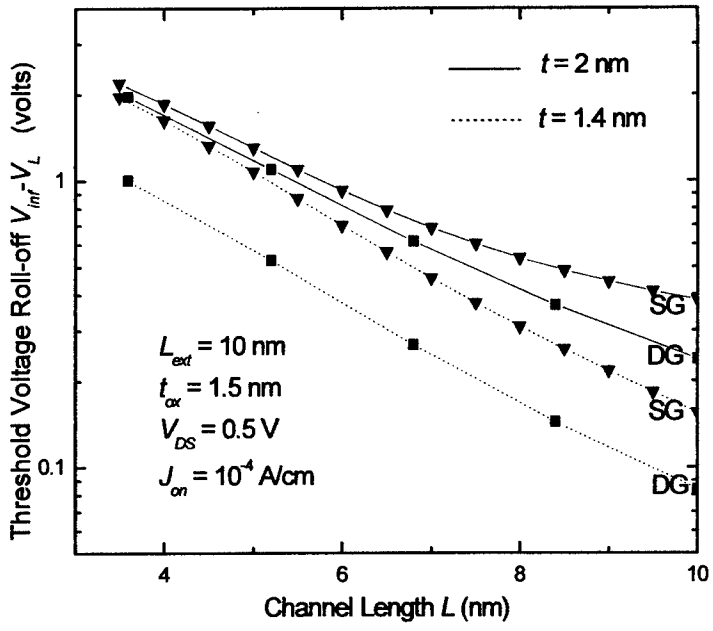

Figure 8. Threshold voltage roll-off as a function of the gate length, for two values of the channel thickness.

4. Conclusions

Our calculations have shown that single-gate SOI MOSFETs differ from double-gate SOI devices (with similar parameters) by twice weaker control of the current by the gate voltage. It is interesting that this relation of the two devices, evident in the depletion range, is also extended, almost exactly, over the ON state range.

By itself, this transconductance (and hence gain) loss might be not very detrimental, taking into account the simpler fabrication technology and also the fact that the internal logic delay of two transistors is comparable (due to the twice larger input capacitance of double-gate transistors). However, single-gate geometry exacerbates two main fundamental problems of the "ultimate" (sub-10-nm) MOSFET scaling: the exponential growth of power consumption and sensitivity to fabrication uncertainties. These factors will play the decisive role in for eventual transfer of the CMOL industry to double-gate SOI devices and then to CMOS/nanodevice hybrids [4, 11].

5. Result Publication

The main results of this study have been published in Ref. 12.

6. Possible Further Work

We see an opportunity to improve our results in two directions:

(i) As pointed recently by V. Trividi and J. Fossum [13], single-gate transistor performance may be improved by making the back gate oxide (BOX) thicker than the front gate oxide. We could readily incorporate this idea into our model and explore scaling limits for this case.

(ii) As follows from our preliminary results [8, 9] , some further improvement of the double-gate MOSFETs may be achieve by the elimination of thin source and drain extensions (Fig. 1), i.e., by transfer to "bulk" electrodes. A quantitative analysis of ultimate scaling of these devices would require using a real 2D Schrödinger equation instead of the quasi-2D equation used earlier.

We would be ready to consider suggestions of funding which would make this work possible.

7. References

1. *International Technology Roadmap for Semiconductors, 2003 Edition, 2004 Update*; available online at http://public.itrs.net/.
2. D. J. Frank, *IBM J. Res. & Devel.* **46**, 235 (2002).
3. G. K. Celler and S. Christoloveanu, *J. Appl. Phys.* **93**, 4955 (2003).
4. K. K. Likharev, "Electronics Below 10 nm", in: J. Greer *et al.* (eds.), *Giga and Nano Challenges in Microelectronics* (Elsevier, Amsterdam, 2003), pp. 27-83.
5. J. Denton and G. Neudeck, *IEEE Trans. on Electron Devices*, **17**, 509 (1996).
6. J. Kedzierski *et al.*, in : *IEDM Tech. Dig.*, p. 437 (2001).
7. S. Harrison *et al.*, in : *IEDM Tech. Dig.*, p. 449 (2003).
8. V. A. Sverdlov, T. J. Walls, and K. K. Likharev, *IEEE Trans. on Electron Devices* **51**, 1926 (2003).
9. T. J. Walls, V. A. Sverdlov, and K. K. Likharev, *Solid State Electronics* **48**, 857 (2004).
10. V. Sverdlov, Y. Naveh, K. Likharev, in: *Proc. ISDRS* (2001), pp. 547-550.
11. K. K. Likharev and D. B. Strukov, "CMOL: Devices, Circuits, and Architectures", to be published as Ch. 16 of collection: G. Cuniberti, K. Richter, and G. Fagas (eds.), *Introduction to Molecular Electronics* (Springer, Berlin, 2005).
12. J. Li , T. J. Walls, and K. K. Likharev, "Nanoscale SOI MOSFETs: In Search for the Best Geometry", in: G. K. Celler (ed.), *SOI Technology and Devices XII* (ECS, Pennington, NJ, 2005), pp. 11-20.
13. V. P. Trividi and J. P. Fossum, *IEEE Electron Device Letters* **26**, 26 (2005).